REPSYS: A robust and distributed incentive scheme for collaborative caching and dissemination in content-centric cellular-based vehicular delay-tolerant networks

Article (Accepted Version)

REPSYS: A robust and distributed incentive scheme for in-network caching and dissemination in Vehicular Delay-Tolerant Networks

Naercio Magaia*†, Zhengguo Sheng†, Paulo Rogério Pereira*, *Senior Member, IEEE*,
Miguel Correia*, *Senior Member, IEEE*
*INESC-ID, Instituto Superior Técnico, Universidade de Lisboa
naercio.magaia@tecnico.ulisboa.pt, prbp@inesc.pt, miguel.p.correia@tecnico.ulisboa.pt
†Department of Engineering and Design, University of Sussex
Z.Sheng@sussex.ac.uk

Abstract

In this article, a robust and distributed incentive scheme for in-network caching and dissemination in cellular-based Vehicular Delay-Tolerant Networks (REPSYS) is proposed. REPSYS is robust because despite taking into account first- and second-hand information, it is resilient against false accusations and praise, and distributed, as the decision to interact with another vehicle depends entirely on each vehicle. The performance evaluation shows that REPSYS is capable, while evaluating each vehicle's participation in the network to classify correctly vehicles in most cases. In addition, it reveals that there are trade-offs in REPSYS, for example, to reduce detection time of vehicles that neither cache nor disseminate other vehicles' data, one may sacrifice the system's resilience against false accusations and praise, or even, by penalizing vehicles that do not disseminate data, one may temporarily isolate vehicles that could contribute to data dissemination.

Index Terms

Caching, Dissemination, Reputation, Trust, Bayesian, Vehicular Delay-Tolerant Network.

I. Introduction

Nowadays, the number of vehicles equipped with devices to provide wireless communication capability that form a vehicular network, also known as *connected vehicles*, have grown significantly. These recently emerged vehicular communication networks are considered as an important technology for improving road safety and building intelligent transportation systems [1].

However, as the number of connected vehicles increases and use cases evolve, the volume of data required for such applications will continue to increase along with the need to minimize latency. The fourth and upcoming generations (5G) mobile cellular networks with broad coverage and high bandwidth are able to provide multimedia content downloading services for the moving vehicles. Nevertheless, with the increase of the services and user demands, these networks will most probably be overloaded and congested especially during peak times and in urban central areas [1]. Therefore, cellular-based vehicular communications will face extreme performance hits in terms of low network bandwidth, missed calls, and unreliable coverage. Caching mobile content at the edge of 5G network(e.g., at the Radio Access Networks - RANs) would help relieve backhaul congestion and meet peak traffic demands with lower service latency [2]. In mobile content sharing, some data items may be more popular and thus needed by a large amount of users. Communication protocols for information transmission between vehicles and roadside unit (RSU) infrastructure equipment, known as vehicles to infrastructures (V2I), as well as between vehicles, known as vehicles to vehicles (V2V), becomes more inevitable for applications of mobile content dissemination. The opportunistic contacts enabled by V2I and V2V communications are capable of providing high bandwidth communication capacity for data transmission, which forms the basis of Vehicular Delay-Tolerant Networks (VDTNs) [3].

In a VDTN, a vehicle will store the data in its buffer, carry it and forwards this data to an appropriate vehicle when a transmission opportunity is available along the vehicle's movement. The latter is known as the *store-carry-and-forward* paradigm. VDTN routing involves the challenging task of finding suitable vehicles to forward data. Service providers can delay or even shift large amounts of data transmissions to the *cellular-based VDTN* (cVDTN), i.e., use in-network caching, by taking advantage of the delay-tolerant nature of some non-real time applications. Although this cVDTN approach may induce tolerable delay for the data dissemination, it helps dealing with the explosive traffic demands and mobile data growth expected nowadays and in the near future.

Even though forwarding schemes have been proposed in literature for VDTNs [4], many challenging and open problems exist in providing efficient data access to moving vehicles, despite the importance of data accessibility in many mobile applications. For example, it is desirable that vehicles are able to find live traffic information that is beneficial to avoid traffic delays. Appropriate network design and/or incentive schemes are hence needed to ensure that data can be promptly accessed by requesters in such cases.

Caching can be used to improve data access performance, i.e., to store data at appropriate cVDTNs locations based on query history, so that queries in the future can be responded with less delay [5]. Although cooperative caching has been studied in literature to allow sharing and coordination among multiple caching vehicles, it is yet an open research problem in cVDTNs due to the lack of persistent network connectivity. However, because of nodes' resource scarcity and the fact of them being controlled by rational entities, they might misbehave. *Vehicle misbehavior*, malicious or selfish, can significantly affect network performance [6]. cVDTN routing as well as caching decision-making becomes much simpler with the use of reputation and trust. *Incentive schemes* [7] can be used to manage and organize decentralized and self-managed systems, hence compensating for the nonexistence of a central or dedicated entity, e.g., for managing reputation and trust.

Reputation-based incentive schemes, hereafter *reputation systems*, are those in which the decision to interact depends on the other vehicle's reputation. A reputation system is composed of three phases: collection of evidence,

cooperation decision and cooperation evaluation [8]. Each vehicle collects reputation information by means of direct experience or special feedback messages (i.e., first-hand information), and by means of other nodes' recommendations (i.e., second-hand information). The collected information is evaluated to decide if the vehicle should cooperate or not based on the other vehicle's reputation. Then, each vehicle evaluates the degree of cooperation with the other vehicle. In a distributed reputation system [7], which is more suitable for cVDTNs as no central authority is available, vehicles' ratings are stored in a distributed fashion and the evaluation of reputation is based on subsets of information (e.g., information provided by neighbor vehicles).

The use of incentive schemes for in-network caching and dissemination in vehicular networks is adequate hence discussed in this article. The REPSYS system is a robust and distributed incentive scheme for in-network caching and dissemination in cVDTNs. Specifically, it is both robust against false ratings and efficient at detecting vehicles' misbehavior (i.e., vehicles that neither forward nor cache or disseminate other vehicles' data). It makes use of all the available information, i.e., first- and second-hand. It is based on a modified Bayesian approach that uses the Beta distribution.

The remainder of this article is as follows. Section II presents the system architecture and elements of the caching system in cellular-based vehicular networks. Section III presents the REPSYS system. In Section IV, the performance evaluation, i.e., the simulation model and results, is presented. Finally, Section V presents conclusions and future work.

II. System architecture and elements of caching system in cellular-based vehicular networks

Figure 1 shows a network topology where vehicles travel around the city roads and the deployed RSUs provide coverage over a certain area. RSUs are placed at the intersections similarly to what is done by current optimal placement algorithms [1]. RSUs are connected through wired links to 5G RANs that are also connected to the content servers in the Internet. Vehicles requiring mobile data such as multimedia newspapers, weather forecasts, movie trailers, etc., send their requests to the content servers via V2I communication links. The requested data is delivered from the content servers to the 5G RAN and from the 5G RAN to the RSUs via the wired links. It is assumed that the wired links provide relatively high bandwidth hence ensuring that the requested data is delivered to RSUs prior to the delay-tolerant dissemination between RSUs and vehicles. RSUs will further disseminate the data to the users in the vehicles that requested it through opportunistic communication that occurs when the vehicle move into the communication coverage of the RSU.

In addition, RSUs and vehicles decide what contents to cache and the RSUs proactively fetch them via backhaul during off-peak times, and transmit the contents to requesters during peak times. By doing so, caching offloads the network traffic during peak times and reduces vehicular users' average delay cost.

V2V communication is used to enhance the benefits of caching. They enable a vehicle to communicate directly with another vehicle in its vicinity. When a vehicle's content request cannot be satisfied by its local cache, any of its neighbors who cached that content can become the content server and transmit the content using V2V communication. The latter improves spectrum utilization, increases network throughput, and reduces average access

delay for vehicles [9]. However, as vehicles are owned or managed by rational entities, they might act selfishly by only caring about the contents of their users, i.e., each vehicle only intends to cache the favorite contents of its users hoping also that its neighbors can cache as many as possible favorite contents of its users.

Now, consider the caching system of Figure 1, where there are 8 contents {1,2,...,8} on the content server. Let each content be of unitary size and the local cache of each vehicle be able to store two contents. Assume that the ranking of content preferences of the users in vehicle V_1 and V_5 are (1,2,...,8), (8,7,6,...,1), respectively. Naturally, V_1 will cache contents {1,2} wishing also that its neighbors will cache contents {3,4,...,8} whereas V_5 will cache contents {8,7} hoping that its neighbors will cache contents {6,5,...,1}. This difference in preferences causes conflict of caching interest among vehicles that may not be settled without intervention. On the other hand, the 5G RAN aims to minimize its traffic load of serving nodes by reducing the backhaul load and the transmission cost. This objective is equivalent to maximizing the chances of V2I and V2V communications among nodes.

The vehicles' selfish nature hence becomes the major obstacle for the 5G RAN to achieve its objective. In a caching system of selfish vehicles, each vehicle cares solely about its own preferences and only cache the contents it likes most. This may cause duplicate caching and underutilization of the storage space for all vehicles. Therefore, the 5G RAN would be overloaded by vehicles' requests and vehicles would suffer from larger delays. It is thus essential for the 5G RAN to introduce incentive schemes into the caching system to motivate vehicles to cache in a way to promote vehicular communication, as described in the next Section.

III. The REPSYS system

The robust and distributed incentive scheme for in-network caching and dissemination in VDTNs (REPSYS) is both robust against false ratings and efficient at detecting vehicles' misbehavior. REPSYS, which is built upon our previous work [8], is robust because despite taking into account all the available information, it is resilient against false accusations and praise, and distributed, as the decision to interact with another vehicle depends entirely on each vehicle. However, differently from [8] that proposed a reputation system to address the routing problem in DTNs, REPSYS is an incentive scheme for cVDTNs that provide incentives to vehicles cache/disseminate data in an opportunistic content sharing vehicular application. REPSYS is based on a modified Bayesian approach that uses the Beta distribution, and uses Bayesian decision theory [10] to classify vehicles.

There are three modules in REPSYS: the reputation module (reputation collection module, reputation evaluation module), the trust module and the decision module (that uses Bayesian classification). Figure 2 shows the block diagram of the REPSYS system.

A. The modified Bayesian approach

Each vehicle considers that there is a given parameter, θ, known as the state of nature such that another vehicle misbehaves with probability θ, and that the outcome is drawn independently at each observation x. Furthermore, each vehicle considers that there is a different θ for every other vehicle. These parameters are unknown, hence modeled according to a prior distribution, $\pi(\theta)$, which is updated as new observations become available.

The beta probability density function $Beta(\theta|\alpha,\beta)$, where $0\leq\theta\leq1$ and the parameters $\alpha,\beta>0$, is used as the prior since it represents probability distributions of binary events (e.g., good or bad). The Bayesian process works as follows. Initially, each vehicle assigns the prior $Beta(\alpha=1,\beta=1)$, that is, the uniform distribution on [0,1], to all the vehicle it meets. The $Beta(1,1)$ prior represents absence of information as there are no observations. When a new observation is made, if a correct behavior is observed then $x=1$; otherwise $x=0$. The prior is updated by summing the current values of both α or β with x.

Due to the network dynamics, a vehicle may change its behavior over time in contrast to the standard Bayesian framework that gives the same weight regardless of time of occurrence of the observation. The fading mechanism allows forgetting gradually old observations [8].

B. Information gathering

Each vehicle is equipped with a pseudo-watchdog component that allows it to monitor the behavior of the other vehicles with whom it interacts. This component can be updated to take into account specific features of vehicular networks. Specifically, if vehicle V_i disseminates data (a query or mobile content) to vehicle V_j, the behavior of V_j is evaluated in terms of two types of evidence, namely: (*i*) if V_j caches data of V_i and, (*ii*) if V_j disseminates V_i's data to another vehicle, say V_k. The former evidence is collected through direct communication between two vehicles (i.e., through experience), meanwhile the latter, is through *Special Feedback Messages* (SFMs). Therefore, V_i waits for an SFM. SFMs can be forwarded using any VDTN routing protocol such as the Epidemic routing protocol [6]. However, other dissemination approaches considering mobility features or predictable trajectories of vehicles could be applied. Two types of SFMs are proposed: (*i*) *type-1* that is created by V_k, which is 2 hops away from V_i (which can be a vehicle that received and cached the data); and (*ii*) *type-2* that is created by the vehicle that requested the data. Each SFM contains the mobile data identifier, the list of vehicles the mobile data traversed and the mobile data digest.

The first-hand information represents the parameters of the Beta distribution assumed by V_i in its Bayesian opinion of V_j's behavior in the caching system. Each vehicle keeps two data structures (records): *cache first-hand information* ($\mathcal{F}_{a_{ij}}$) for cached data and *disseminate first-hand information* ($\mathcal{F}_{f_{ij}}$) for disseminated data. For each record there are two counters: α and β. Cache and disseminate first-hand information are given by $\mathcal{F}_{x_{ij}}=(\alpha,\beta)_x$, where $x \in \{c,d\}$, and they are updated to identify *attacks' signature* as follows:

- α is incremented if a good behavior is observed when:
  - V_j caches data of other vehicles, e.g., V_i. However, only caching others' data may not be optimal for the caching system besides being an indicator of a *black-hole attack*. Therefore, it is also necessary to ensure that V_j disseminates data that it caches if the data was not requested by it; or
  - V_i receives an SFM from V_k because of the data V_i disseminates to V_j.
- β is incremented if a misbehavior is observed when:
  - vehicle V_j not being the vehicle that requested the data disseminated by V_i, does not disseminate this data (no SFM was received neither did the data expire); or

- V_j does not cache data of other vehicles, e.g., V_i, which may be an indication that V_j is performing a *lying attack*. V_j can only refuse to cache data disseminated to it, if it already has the data in its local cache or by proving that the data will be discarded to make space for other more requested data.

Since only using first-hand information may not be cost-effective, reputation systems that exclusively rely on it might have higher detection times in comparison with other approaches that also use second-hand information $\mathcal{S}_{ij} = (\alpha, \beta)_{\mathcal{S}}$. A faster convergence of a reputation system is more likely as more information is considered by each vehicle. Second-hand information corresponds to first-hand information published by other vehicles.

C. Reputation rating

The reputation module is responsible for managing reputation ratings. A reputation rating $\mathcal{R}_{ij}$ is updated (*i*) when first-hand information is updated, and (*ii*) when received second-hand information is considered valid to be incorporated.

If cache and disseminate first-hand information that are kept by each vehicle are available, they are combined to form a unique first-hand information, hereafter called first-hand information $\mathcal{F}_{ij}=(\alpha,\beta)_{\mathcal{F}}$. The first-hand information rating corresponds to the expectation of $Beta(\alpha,\beta)_{\mathcal{F}}$. When first-hand information is updated, an exponential weighted moving average (EWMA) is used to update the reputation rating therefore allowing for reputation fading. Since classical EWMA averages do not take into account time, at the end of a given time interval, first-hand information is updated by means the fading mechanism. When received second-hand information is considered valid to be incorporated, linear opinion pooling [11] is used for its integration.

Assume two vehicles V_i and V_k where V_i has its opinion on how honest V_k is as an actor in the reputation system and V_k collects first-hand information about V_j. A *recommendation* then consists in combining V_i's opinion about V_k with V_k's opinion about V_j in order for V_i to get its opinion about V_j.

Any vehicle V_k's recommendations towards V_j are synthetized and integrated using the same moving average process used to update the reputation rating, thus making the system resilient against false praise and accusation.

D. Trust rating

The trust module is responsible for managing trust ratings. The trust record has the form $\mathcal{T}_{ij}=(\alpha,\beta)_{\mathcal{T}}$. $Beta(\alpha,\beta)_{\mathcal{T}}$ represents the parameters of the Beta distribution assumed by vehicle V_i in its opinion about how honest V_j is as an actor in the reputation system. When V_i receives first-hand information from some V_k about V_j, an update is performed.

Prior to incorporating the second-hand information, a deviation test is executed. The deviation test allows comparing if vehicles V_i and V_k have similar opinions about V_j by comparing the absolute difference of the accumulated rating and the received one with the deviation threshold. On the one hand, it is used to update the trust rating V_i has of V_k, and on the other hand and in addition to the latter, it is also used to decide whether to update the reputation rating V_i has on V_j.

Similarly to first-hand information rating, the trust rating corresponds to the expectation of $Beta(\alpha,\beta)_{\mathcal{T}}$.

E. Bayesian classification

The decision module is responsible for: (*i*) classifying vehicles based on their behavior and (*ii*) taking caching and dissemination decisions. In classification problems, Θ is discrete and the goal is to estimate θ given an observation x. To address the in-network caching and dissemination problem in cVDTNs, the vehicle's behavior classification problem is considered.

Let

- $\theta \in \Theta$= $\{\theta_0$=NORMAL,θ_1=MISBEHAVING$\}$ unknown state of nature.
- $X \in \mathcal{X}$ be a random variable with $\{f(x|\theta), x \in X\}$
- $\pi(\theta) > 0$ and $\sum_{\theta \in \Theta} \pi(\theta) = 1$ be the prior probability mass function
- $a \in \mathcal{A}$= $\{a_0$=CACHE_DISSEMINATE,a_1=DO_NOT_CACHE_DISSEMINATE$\}$ be the allowed decision or action.
- The "0/1" loss function be used for classification. It assigns zero cost to any correct decision and unit cost to any wrong decision.
- $\mathcal{D}$ be the set of allowed decision rules. A decision rule ($\delta(x)$) specifies how actions or decisions are chosen given x.
- $L(\theta,a)$ be the loss function. It quantifies the consequences of the decisions.

The optimal Bayesian decision $\left(\delta_{\text{Bayes}}(x)\right)$ is equal to θ_0 if the ratio between $f(x|\theta_0)$ and $f(x|\theta_1)$, also known as the likelihood ratio, is greater or equal to the ratio between $\pi(\theta_0)$ and $\pi(\theta_1)$, also known as the decision threshold. Otherwise, $\delta_{\text{Bayes}}(x)$ is equal to θ_1. The likelihood function is given by the Bernoulli distribution.

In the beginning, if the only information available is the conditional probability density function of the observation given the true θ, the maximum likelihood decision criterion (δ_{ML}) [12] is used. δ_{ML} is equal to θ_0 if the likelihood criterion is greater or equal to 1. Otherwise, δ_{ML} is equal to θ_1.

In the vehicle's behavior classification problem, after each interaction between two vehicles, the sender updates the reputation rating of the other vehicle based on the result of this interaction. Each vehicle clusters the other vehicles with whom it interacted in two groups: normal vehicles, if $\mathcal{R}_{\text{ij}} \geq 1/2$, and misbehaving vehicles, if $\mathcal{R}_{\text{ij}} < 1/2$. The prior probabilities $\pi(\cdot)$ of these clusters, which allow determining the decision threshold, are coefficients of the convex combination of the number of vehicles in these clusters. The optimal Bayesian decision is computed as previously explained given the prior probabilities. However, if a correct behavior is observed and $\pi(\theta_1) > \pi(\theta_0)$, one may incur in false positives, i.e., a misclassification, while using the optimal Bayesian decision criterion, because of the higher weight of the decision threshold in comparison to the likelihood ratio.

A modified optimal Bayesian decision was used as the workaround. It consists in finding attenuation parameters α and β of the *posterior mean Bayesian estimator* [13] and computing an attenuated decision threshold. For the minimum possible case, i.e., one correct behavior being observed and two clusters, one with 2 misbehaving vehicles and the other with 1 normal vehicle, the likelihood ratio is $4/3$. For this case and with the Bayesian attenuation parameters α=β=2, the decision threshold is equal to the likelihood ratio. If instead the *maximum a posteriori Bayesian estimator* [13] was used, the decision threshold would be greater than the likelihood ratio which would

lead to misclassification.

IV. Performance evaluation

This section presents the simulation model and results regarding the performance evaluation of REPSYS.

A. The simulation model

REPSYS was implemented on the Opportunistic Network Environment (ONE) simulator [14]. The simulation model consisted of a synthetic mobility model (SMM) and a real mobility trace (RMT). The simulation time was 7 days with an update interval of 1.0 s. The deviation threshold value was set to $1/6$. The latter value means that only second-hand information rating whose difference to the first-hand information rating stored by the vehicle is less of or equal to $1/6$ will be incorporated. The vehicles misbehavior considered for evaluation was the black-hole attack. It was considered that misbehaving vehicles were also colluding, that is, they increased α of misbehaving vehicles and β of normal vehicles. The effects of vehicles' misbehavior was examined considering that vehicles were using an Epidemic approach to disseminate data. The percentage of vehicles that performed black-hole attacks varied from 20% to 80% with increments of 20%. It was considered that queries were generated randomly every 60 to 120 seconds. It also assumed that queries propagated almost instantaneously to the content server or node containing the data. The data size varied from 50KB to 500KB.

SMM consisted of a network with 150 vehicles and it was configured similarly to [8]. The RMT considered was taxicabs in Rome (TR) [15]. TR contains Global Positioning System (GPS) coordinates of approximately 320 taxicabs collected over 30 days in Rome, Italy. The simulation duration and number of vehicles of RMT were reduced to 7 days and 304 vehicles, respectively. All vehicles had a buffer size of 10 MB for cVDTN traffic. By considering such small buffer sizes, the caching problem became more challenging. The TTL attribute of each content was 24 h.

B. Simulation results

The evaluation of the performance of REPSYS consisted in appraising the reputation and trust modules, similarly to previous work [8]. Additionally, Bayesian classification at the decision module was also evaluated. For each setting, i.e., protocol-percentage pair, thirty independent simulations using different query message generation seeds were conducted, and the results averaged, for statistical confidence.

The following metrics were considered for the evaluation of REPSYS:

- *Detection time of misbehaving vehicles* corresponds to the simulation time that took all normal vehicles to correctly classify all misbehaving vehicles they came in contact with, starting at the detection instant of the first misclassification.
- *Robustness* against false accusations (false negatives) and false praise (false positives). The following metrics were defined:
  - Vehicle's Behavior False Positives Ratio (VBFPR) is the number of misbehaving vehicles with normal vehicle's behavior classification, i.e., classified as CACHE_DISSEMINATE, over all vehicles classified.

- Vehicle's Behavior False Negatives Ratio (VBFNR) is the number of normal vehicles with bad vehicle's behavior classification, i.e., classified as DO_NOT_CACHE_DISSEMINATE, over all vehicles classified.

1) Detection time of misbehaving vehicles: Figures 3 presents the time necessary for each normal vehicle to classify correctly all misbehaving vehicles it met as DO_NOT_CACHE_DISSEMINATE in both scenarios.

REPSYS used all the available information to infer the behavior of each vehicle with which it interacts. Still, in some cases, even though there were many evidence that a given vehicle cached many data and did not disseminate any, one could not say for sure that this vehicle was misbehaving since some normal vehicles also presented a similar behavior.

In both scenarios, REPSYS penalizes vehicles that only cached but did not disseminate data given that evidence that these data were not disseminated expired. Even if a small penalization was given, misbehaving vehicles performing black-hole attacks were detected. However, normal vehicles that behaved similarly to misbehaving vehicles were also isolated from the network, although temporarily, because of the fading mechanism or if they started disseminating data.

For SFM *type-2*, since an evidence has, by default, the same TTL of a data that originated it, there is a tradeoff between the TTL and the detection time. If the goal is for REPSYS to converge sooner (i.e., to have a small detection time) then the TTL should not be too high. Otherwise, SFMs might not have enough time to be effectively propagated over the network, which would increase the number of misclassifications as a consequence of a too small TTL. Nevertheless, REPSYS took more time to detect an increasing percentage of vehicles performing black-hole attacks mainly because of disseminate first-hand information.

Figure 3 also shows that vehicles in RMT took more time to start detecting and classifying correctly the misbehaving vehicles they met. Specifically, the vehicles took 4.57 days to start detecting and correctly classifying 20% of misbehaving vehicles. On the other hand, the vehicles in SMM took 2.46 hours to do the same.

2) Robustness: In Figures 4, two metrics were considered to measure REPSYS's robustness against false accusations and praise for the black-hole attack in both scenarios.

The use of second-hand information may lead to false accusations and praise, but even with the optimal Bayesian decision criterion, it did not have any influence on the robustness metrics considered. There are two reasons for that: (*i*) the bootstrapping of the trust module and (*ii*) the tolerance to vehicles that failed the deviation test. Recall that the deviation test allows each vehicle to synthesize first-hand information received from other vehicles (i.e., collected and accumulated using EWMA). By comparing the received information with the accumulated one in each vehicle, the probabilities of false praise and accusations were small. But then again, as on cVDTNs many vehicles get isolated, it was noticed that some vehicles failed the deviation test because of stale accumulated information. Consequently, each vehicle should tolerate failures to the deviation test up to a given number of times. The combination of these two techniques allowed the trust module to presents zero false positives and negatives in most of the cases.

Additionally, there is also a tradeoff between false positives and negatives. By attempting to isolate misbehaving vehicles (that is, to reduce the false positives ratio), normal vehicles that up to a given instant only cached data will be misclassified as DO_NOT_CACHE_DISSEMINATE, therefore increasing the ratio of false negatives.

V. Conclusions and future work

In this article, a robust and distributed incentive scheme for in-network caching and dissemination in VDTNs (REPSYS) was presented. REPSYS takes into account all the available information and uses Bayesian decision theory to classify vehicles.

The emerging latency requirement of the future 5G vehicular networks rely on the cooperative behavior of the vehicles. REPSYS can play an important role by providing incentives to vehicles start sharing their resources by caching and/or disseminating other vehicles' data thus reducing latency to the users that requested the data. The performance evaluation shows that the system is able to classify correctly vehicles in most cases. In addition, there are tradeoffs in this system. For instance, if the evidence's TTL is too high, the reputation system will take more time to converge as the detection time increases.

As future work, the following research challenges have been identified: (*i*) the evaluation of REPSYS with other caching and dissemination approaches, and (*ii*) the use of more elaborate attacker scenarios such as mixing lying and black-holes attacks with different bootstrapping periods.

Acknowledgment

This research was partially supported by Fundação Calouste Gulbenkian and by national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UID/CEC/50021/2013. It was also sponsored by The Engineering, and Physical Sciences Research Council (EPSRC) (EP/P025862/1), Royal Society-Newton Mobility Grant (IE160920).

References

[1] Y. Li, D. Jin, P. Hui, and S. Chen, "Contact-aware data replication in roadside unit aided vehicular delay tolerant networks," *IEEE Transactions on Mobile Computing*, vol. 15, no. 2, pp. 306–321, Feb 2016.

[2] Y. Fadlallah, A. M. Tulino, D. Barone, G. Vettigli, J. Llorca, and J. M. Gorce, "Coding for caching in 5g networks," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 106–113, February 2017.

[3] P. R. Pereira, A. Casaca, J. J. P. C. Rodrigues, V. N. G. J. Soares, J. Triay, and C. Cervello-Pastor, "From delay-tolerant networks to vehicular delay-tolerant networks," *IEEE Communications Surveys Tutorials*, vol. 14, no. 4, pp. 1166–1182, Fourth 2012.

[4] N. Benamar, K. D. Singh, M. Benamar, D. E. Ouadghiri, and J.-M. Bonnin, "Routing protocols in vehicular delay tolerant networks: A comprehensive survey," *Computer Communications*, vol. 48, pp. 141 – 158, 2014.

[5] K. Pentikousis, B. Ohlman, D. Corujo, G. Boggia, G. Tyson, E. Davies, A. Molinaro, and S. Eum, "Information-Centric Networking: Baseline Scenarios," *RFC*, no. 7476, 2015. [Online]. Available: https://tools.ietf.org/html/rfc7476

[6] N. Magaia, P. R. Pereira, and M. P. Correia, "Selfish and malicious behavior in Delay-Tolerant Networks," in *Future Network and Mobile Summit (FutureNetworkSummit)*, 2013, pp. 1–10.

[7] N. Magaia, P. Pereira, and M. P. Correia, "Security in Delay-Tolerant Mobile Cyber Physical Applications," in *Cyber-Physical Systems: From Theory to Practice*, D. B. Rawat, J. J. P. C. Rodrigues, and I. Stojmenovic, Eds. CRC Press, 2015, ch. 15, pp. 373–394. [Online]. Available: http://www.crcnetbase.com/doi/abs/10.1201/b19290-22

[8] N. Magaia, P. R. Pereira, and M. Correia, "REPSYS: A Robust and Distributed Reputation System for Delay-Tolerant Networks," in *20th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM)*, Nov 2017, pp. 1–5.

[9] Z. Chen, Y. Liu, B. Zhou, and M. Tao, "Caching incentive design in wireless d2d networks: A stackelberg game approach," in *2016 IEEE International Conference on Communications (ICC)*, May 2016, pp. 1–6.

[10] J. O. Berger, *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media, 2013.

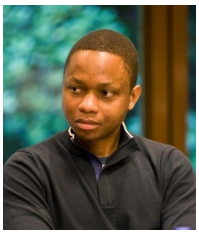
[11] F. Dietrich and C. List, "Probabilistic Opinion Pooling," in *The Oxford Handbook of Probability and Philosophy*, A. Hájek and C. Hitchcock, Eds. Oxford University Press, 2016, ch. 25, p. 832.

[12] J. L. Melsa and D. L. Cohn, *Decision and estimation theory*. McGraw-Hill, 1978.

[13] M. A. T. Figueiredo, "Lecture notes on Bayesian estimation and classification," Instituto de Telecomunicações, Instituto Superior Técnico, Lisboa, Tech. Rep. October, 2004.

[14] A. Keränen, J. Ott, and T. Kärkkäinen, "The ONE simulator for DTN protocol evaluation," in *Proceedings of the 2nd international conference on simulation tools and techniques*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2009, p. 55.

[15] L. Bracciale, M. Bonola, P. Loreti, G. Bianchi, R. Amici, and A. Rabuffi, "CRAWDAD dataset roma/taxi (v. 2014-07-17)," Downloaded from http://crawdad.org/roma/taxi/20140717, Jul. 2014.



Naercio Magaia received his PhD with distinction in Electrical and Computer Engineering at Instituto Superior Técnico (IST), Universidade de Lisboa (ULisboa). He holds a B.Sc. in Electrical Engineering from Eduardo Mondlane University, and a M.Sc. in Communication Networks Enginnering from IST, ULisboa. His current research interests cover vehicular delay-tolerant networks, network security, edge computing and multi-objective optimization.

Zhengguo Sheng has been a lecturer in the Department of Engineering and Design at University of Sussex since 2015. He received his Ph.D. and M.S. with distinction at Imperial College London in 2011 and 2007, respectively, and his B.Sc. from the University of Electronic Science and Technology of China (UESTC) in 2006. From 2013 to 2014, he was a research associate in the Department of Electrical and Computer Engineering at University of British Columbia (UBC), Canada. From 2011 to 2013, he was with France Telecom Orange Labs as the senior researcher and project manager in M2M/IoT. During 2009, he also worked as a research intern with IBM T. J. Watson Research Center, USA, and U.S. Army Research Labs. His current research interests cover Internet-of-Things (IoT), connected vehicles, and cloud/edge computing.

Paulo Rogério Pereira (S'97, M'04, SM'15) received his Ph.D. in Electrical and Computer Science Engineering from Instituto Superior Técnico, University of Lisbon (IST/UL), Portugal, in 2003. He is an assistant professor of computer networks at IST/UL and a senior researcher at INESC-ID. He has participated in the IST European projects EuroNGI, EuroFGI, EuroNF, UbiSec&Sens, WSAN4CIP and E-Balance. His research interests include IP wireless sensor networks, delay-tolerant networks, quality of service and network management.

Miguel Correia is an Associate Professor at Instituto Superior Técnico (IST) of the Universidade de Lisboa (ULisboa), in Lisboa, Portugal. He is a researcher at INESC-ID in the Distributed Systems Group (GSD). He is currently the coordinator of the Degree in Computer Engineering at IST (alameda campus). He has a PhD in Computer Science from the University of Lisboa Faculty of Sciences. He has been involved in several international and national research projects related to intrusion tolerance and security, including the PCAS, TCLOUDS, MAFTIA and CRUTIAL European projects, and the ReSIST network of excellence. He has more than 100 publications. His main research interests are: security, intrusion tolerance, distributed systems, distributed algorithms, computer networks, cloud computing, and critical infrastructure protection.

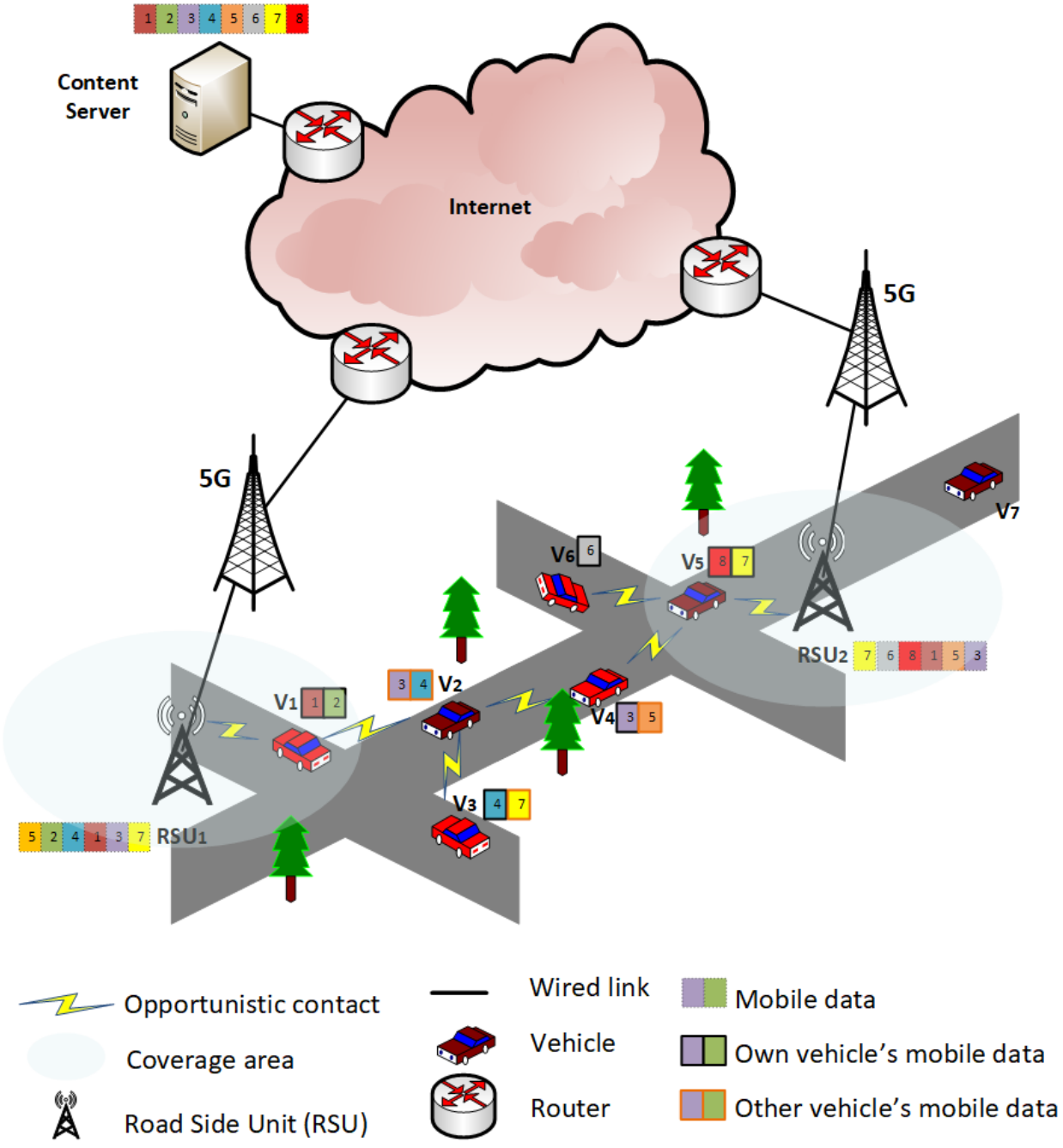


Fig. 1. Illustration of the caching system integrating cellular network and opportunistic communications.

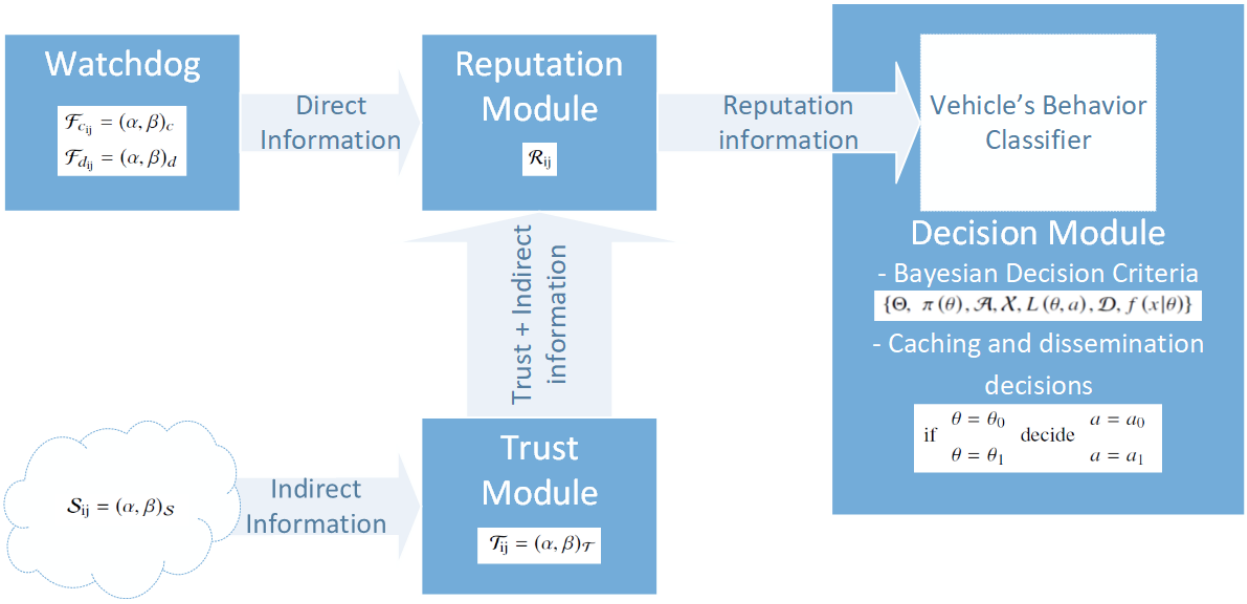
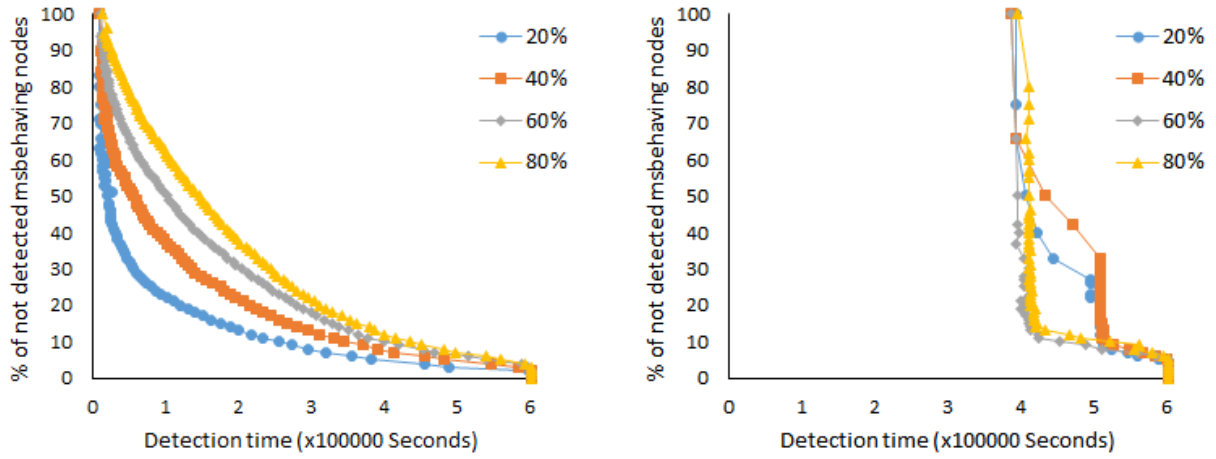
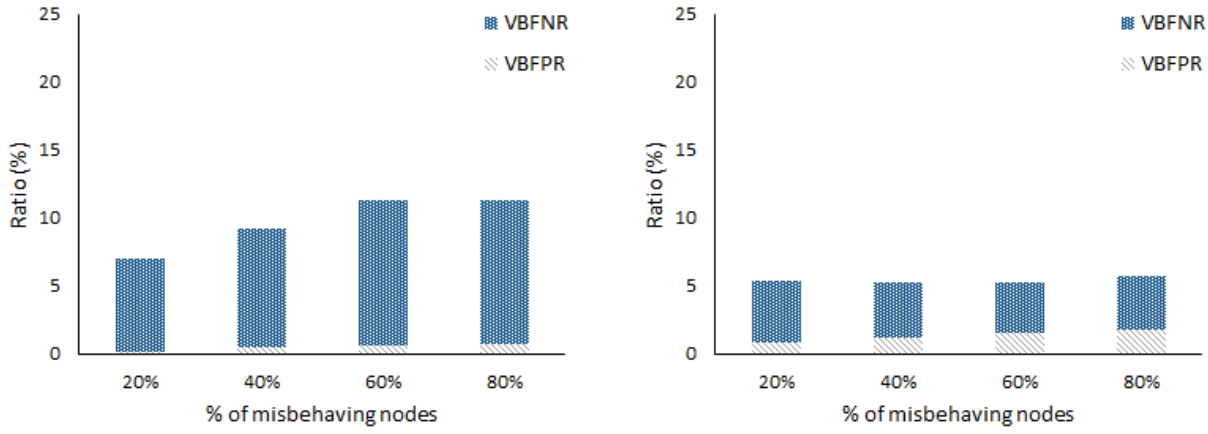


Fig. 2. A block diagram of the REPSYS system



(a) SMM (b) RMT

Fig. 3. The time necessary to correctly classify misbehaving vehicles as DO_NOT_CACHE_DISSEMINATE for 20, 40, 60 and 80% of black-hole vehicles for SMM and RMT scnearios



(a) SMM (b) RMT

Fig. 4. Vehicle's behavior false positives and negatives ratios for 20, 40, 60 and 80% of black-hole vehicles for SMM and RMT scenarios